

Data Storage Apparatus

Field of the Invention

The present invention relates to data storage apparatus, particularly but not exclusively apparatus arranged to provide redundancy in a storage enclosure containing multiple Serial ATA disk drives. In this connection it is to be understood that a storage enclosure is a rack containing multiple disk drives for the bulk storage of data.

Background of the Invention

Serial ATA is an emerging disk connection standard, designed to replace the ATA connection inside standard personal computers (PCs). ATA is also known as Integrated Drive Electronics (IDE).

Fig. 1 shows a conventional connection between a host 10 (e.g. PC, RAID controller or disk drive controller) and a Serial ATA disk drive 12. There are two serial data paths, one sending commands and data from the host to the drive (transmit path 14), and one sending commands and data in the reverse direction (receive path 16). Each data path comprises a differential pair of signals.

It will be noted that if the host or disk controller 10 fails then it is not possible to access the data on the disk drive 12.

Unlike Fibre Channel disk drives, serial ATA drives do not have redundant ports as they only have one port to connect to the host or disk controller. Since there is

only one port and since serial ATA connections are point to point links there is no option for redundant paths to the disk drive.

5 Some solutions to this problem include:

- Making a Serial ATA disk drive with two ports; however, this requires an expensive customised disk with a non-standard connector; or
- 10 • Mirroring the data on all Serial ATA disk drives so that there is a redundant path to the data rather than the disk; however, this requires twice the number of disk drives.

15 **Disclosure of the Invention**

Accordingly, the present invention provides data storage apparatus comprising at least one disk drive of a kind having a single port for the input and output of serial data, at least two disk drive controllers each
20 having data transmit and receive paths connected in common to the single port, and means for switching control to either controller if the other should fail.

Preferably the transmit paths of the controllers are
25 connected to the disk drive via respective inputs of a first changeover switching device.

In one embodiment the transmit path of a first controller is connected to the respective input of the
30 first changeover switching device via an input of a second changeover switching device, and wherein the receive path of the second controller is connected to

the second controller via another input of the second changeover switching device.

Alternatively, the receive paths of the controllers are
5 connected to the respective controllers via a buffer.

In another embodiment the switching is effected solely by software in the controllers.

10 The disk drive(s) are preferably serial ATA device(s).

In one embodiment a redundant path is provided by the addition of some low cost chips housed on a circuit board in a Serial ATA storage enclosure. Preferably,
15 the circuit board comprises a printed circuit board arranged to attach to the rear of the disk drive.

Potentially all Serial ATA storage enclosures could use this invention. However, the invention is not
20 restricted to use within a storage enclosure; for example, it could be used within a server. Modern servers can have several disk drives (e.g. 8) within the server box. The invention can also be used with individual disk drives.

25

Brief Description of the Drawings

Embodiments of the invention will now be described with reference to the accompanying drawings, in which:

30 Fig. 1 shows a conventional connection between a host and a Serial ATA disk drive;

Fig. 2 shows a first embodiment of redundant connection according to the invention;

Fig. 3 shows a second embodiment of redundant
5 connection according to the invention;

Fig. 4 shows a third embodiment of redundant connection according to the invention; and

10 Fig. 5 illustrates a physical implementation of the first embodiment of the invention.

Description of the Preferred Embodiments

When describing embodiments of the invention, it will
15 be noted that, for clarity, all the figures only show a single disk drive, while in reality there would be many drives in the case of storage enclosures.

Referring now to the embodiment of Fig. 2, two disk
20 drive controllers 10a, 10b are connected in common to the single port of the Serial ATA disk drive 12 (the disk drive controllers 10a, 10b may be dedicated controllers or host PCs which perform a disk drive control function among other functions). These
25 controllers may be fixed or removable. Their transmit paths are indicated by the arrows pointing towards the disk drive 12 and the receive paths are indicated by the arrows pointing away from the disk drive. If one controller fails then the data on the disk can be
30 accessed by the other. This is effected as follows.

The transmit path of the first controller 10a is connected to the disk drive 12 via a respective input

of a multiplexer Mux A acting as a changeover switch. Thus the transmit path of the first controller 10a is 14a/14. The transmit path of the second controller 10b is via a respective input of a second multiplexer Mux B, also acting as a changeover switch, and the other input of the first multiplexer Mux A. The transmit path of the second controller is thus 14b/14c/14. The receive path of the first controller is via the other input of the second multiplexer Mux B; thus its receive path is 16/16c/14c/16a. Finally, the receive path of the second controller 10b is direct, i.e. 16/16b.

When the controller 10a is active multiplexer Mux A is configured to send data from Host A to the disk drive 12, and multiplexer Mux B is configured to send data from the drive back to Host A. When Host B is active multiplexer Mux B is configured to send data from Host B to multiplexer Mux A and multiplexer Mux A is configured to send the data to the drive. Thus a malfunctioning host can be bypassed completely.

Controllers 10A and 10B communicate in any suitable manner (e.g. RS232, Ethernet) to agree which of them is going to communicate with the disk drive at any given time. A redundant scheme exists between the controllers to allow them to operate in a highly available manner, including, for example, the so-called STOMITH (Shoot The Other Machine In The Head) method.

The two controllers jointly control Mux A and Mux B in such a manner that if controller 10a is not present or is not working properly then multiplexer Mux A and multiplexer Mux B are switched so as to put controller

10b in control of the drive, and vice versa. This can be achieved using signals from the controllers in conjunction with pull-up or pull-down resistors. (Note that when controller 10a is using the STOMITH feature to reset controller 10b then the latter's Mux control lines must be set into a benign state when in reset.)

Referring now to the embodiment of Fig. 3, in this case the data from the disk drive 12 is fed to both controllers via a buffer Buffer A. In certain controlled situations the buffer may not be necessary but is shown here for completeness. Both controllers receive the data from the disk drive but the inactive one ignores it. It should be noted that in this case, one multiplexer, Mux A, is still required to connect the correct controller to the disk drive 12 on the transmit path.

Referring now to the embodiment of Fig. 4, in this case the transmit path of two controllers are connected together and jointly connected to the disk drive 12. System software controls access to the disk drive and prevents both controllers attempting to transmit at the same time. The inactive controller becomes totally passive. Again, in certain controlled situations the buffer may not be necessary but is shown here for completeness. In such a case the Fig. 4 embodiment may not require any extra hardware components and thus essentially consists of wiring together what's already there.

The embodiments outlined in Figs. 2, 3 and 4 could be implemented in several ways; for example, the

multiplexers and buffers could be on the controllers, on the backplane of the disk drive or on a small board between the backplane and the disk drive. (Note that the Fig. 4 embodiment may not need any boards at all)

5

Fig. 5 shows the second embodiment as an example. In Fig. 5 a single disk drive 12 is shown, together with a small PCB 20 which is mounted on the rear of it. When the disk drive is removed the PCB comes with it, thus
10 allowing the backplane 22 to be a completely passive board with no active devices on it, thus making it more reliable and easier to service. In this example the controllers 10a, 10b are shown as removable boards. The dashed line 24 indicates the communication between the
15 two controllers, as mentioned above.

The invention is not limited to the embodiments described herein which may be modified or varied without departing from the scope of the invention.